AD-A039 951 NAVAL POSTGRADUATE SCHOOL MONTEREY CALIF F/G 12/1

DETERMINATION OF HYPOTHESIS TESTABILITY IN LINEAR STATISTICAL M--ETC(U)

MAR 77 W H WALLS

UNCLASSIFIED NL

1 OF 1

AD A039951

END

DATE FILMED

6-77

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

DETERMINATION OF HYPOTHESIS TESTABILITY IN LINEAR STATISTICAL MODELS

BY

William Hammond Walls

March 1977

Thesis Advisor: D. R. Barr

Approved for public release; distribution unlimited.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) DETERMINATION OF HYPOTHESIS TESTABILITY IN LINEAR STATISTICAL MODELS.		5. TYPE OF REPORT & PERIOD COVERED (9) Master's Thesis; ~~March~~ 1977
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) William Hammond Walls		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE (11) March 1977
		13. NUMBER OF PAGES 50
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Naval Postgraduate School Monterey, California 93940 (12) 50p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Analysis of variance, linear model, hypothesis testing, estimability, statistics, testability, generalized inverse, design of experiments		

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Analysts conducting experiments must frequently deal with situations in which data is incomplete or missing. This creates problems that can seriously affect classical hypothesis testing by introducing extraneous terms into the hypothesis in a complicated way. A technique exists that allows an analyst to determine precisely which experimental terms are actually present in a proposed hypothesis and what that hypothesis would actually be testing if employed. This paper examines the mathematics underlying the technique and → next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
(Page 1) S/N 0102-014-6601 |

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251450

applies the theory to a widely used data analysis computer package. A computer program is presented to facilitate implementation of the method.

ACCESSION for		
NTIS	White Section	☒
DDC	Buff Section	☐
UNANNOUNCED		☐
JUSTIFICATION		
BY		
DISTRIBUTION/AVAILABILITY CODES		
DIST.	AVAIL. and/or SPECIAL	
A		

Approved for public release; distribution unlimited

DETERMINATION OF HYPOTHESIS TESTABILITY IN LINEAR STATISTICAL MODELS

by

William Hammond Walls
Lieutenant-Commander, United States Navy
B. S., United States Naval Academy, 1966

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL

March 1977

Author: William H. Walls

Approved by: Donald R Barr
Thesis Advisor

F R Richards
Second Reader

Michael G Sovereign
Chairman, Department of Operations Research

D A Schrady
Dean of Information and Policy Sciences

ABSTRACT

Analysts conducting experiments must frequently deal with situations in which data is incomplete or missing. This creates problems that can seriously affect classical hypothesis testing by introducing extraneous terms into the hypothesis in a complicated way. A technique exists that allows an analyst to determine precisely which experimental terms are actually present in a proposed hypothesis and what that hypothesis would actually be testing if employed. This paper examines the mathematics underlying the technique and applies the theory to a widely used data analysis computer package. A computer program is presented to facilitate implementation of the method.

TABLE OF CONTENTS

I. EXECUTIVE SUMMARY

A. PROBLEM

At times even the most carefully designed and executed experiments can be plagued with aborted tests or missing data. Such unbalance in the data can have a significant impact upon the mathematical structure of analytic techniques used in analysis of variance. In addition to increasing the complexity of computations, unbalanced design can also seriously affect hypothesis testing. Because of lack of balance, hypotheses purporting to test the influence of a main effect, for example, may be hopelessly confounded with interaction terms. Blindly "testing" such confounded hypotheses without an appreciation of the level of pollution from extraneous terms can lead to serious error in interpreting results. It is desirable to find a general procedure for use with analysis of variance that can determine exactly what a proposed hypothesis is testing in terms of the main effects and interactions.

B. APPROACH

Because of its mathematical power and notational simplicity, the matrix form of the linear model $Y = Xb + e$ is used in deriving a solution to the problem. The linear model leads to the "normal equations" $X'Xb = X'Y$. Since $X'X$

is in general not of full rank, any solution (b^0) for b is not unique. Further $(X'X)^{-1}$ does not exist; one must turn to the concept of a generalized inverse G of $X'X$. It can be shown that testing a hypothesis H_0: $q'b = m$ involves expressing the hypothesis as a linear function $q'GX'X$ of the generalized inverse (G) and $X'X$. While determination of $q'GX'X$ is frequently a non-trivial manual calculation, it can be handled easily on a computer.

C. SOLUTION

If an analyst needs to test a particular hypothesis it is possible that additional, undesired terms may be polluting the hypothesis to such a degree that his interpretation of test results may be completely invalid. By computing the value of $q'GX'X$ he will be able to determine precisely what his proposed hypothesis is actually testing.

D. CONCLUSICNS

Recognizing that an unbalanced design can lead to difficulty in interpreting traditional tests of hypotheses, it is concluded that:

1. it is mathematically possible to determine the exact nature of a proposed hypothesis, and

2. such a determination is feasible using a computer.

II. MATHEMATICAL JUSTIFICATION

A. GENERALIZED INVERSE

A generalized inverse of a matrix A is defined to be any matrix G satisfying

$$AGA = A.$$

It can be shown that, for a given matrix A, G is in general not unique [Searle, 1971].

B. SOLUTION OF CONSISTENT LINEAR EQUATIONS

The system of linear equations $AX = Y$ is consistent if any linear relationships existing among the rows of A also exist among the corresponding elements of Y. Since linear equations have a solution if and only if they are consistent, the procedures outlined below are confined to systems of consistent linear equations.

The following theorems from Searle [1] are stated without proof in order to develop solution procedures for consistent equations.

Theorem 1. Consistent equations $AX = Y$ have a solution $X = GY$ if and only if $AGA = A$.

Theorem 2. If A has q columns and if G is a generalized inverse of A, then the consistent equations $AX = Y$ have the

solution

$$X^0 = GY + (GA - I)Z$$

where Z is any arbitrary vector of order q. The notation indicates that X^0, which satisfies $AX = Y$, is a solution and not the general vector of unknowns X.

Theorem 3. For the consistent equations $AX = Y$, all solutions are, for any specific G, generated by $X^0 = GY + (GA - I)Z$, for arbitrary Z. That is, one need derive only one generalized inverse of A in order to be able to develop all solutions to the system $AX = Y$.

C. THE SPECIAL CASE OF SYMMETRIC MATRICES

The linear model used, inter alia, in analysis of variance involves the system of consistent linear equations

$$X'Xb = X'Y$$

that are solved for b. It is therefore worthwhile to consider the special case of the symmetric matrix $X'X$ in some detail. The following development is from Searle [1].

Lemma 1. $X'X = 0$ implies $X = 0$.
Proof: This is true because if $X'X = 0$, the sums of squares of the elements of each row equal zero, hence must be zero.

Lemma 2. $PX'X = QX'X$ implies that $PX' = QX'$.
Proof: Apply Lemma 1 to the identity

$$(PX'X-QX'X)(P-Q)' = (PX'-QX')(PX'-QX')' = 0.$$

That is, $(PX'-QX')(PX'-QX')' = 0$ implies that $(PX'-QX') = 0$ which implies that $PX' = QX'$.

Theorem 4. When G is a generalized inverse of $X'X$, then

i. G' is also a generalized inverse of $X'X$;

ii. $XGX'X = X$ (i.e., GX' is a generalized inverse of X);

iii. XGX' is invariant to G.

Proof:

(i) By definition

$$X'XGX'X = X'X$$

transposing $X'XG'X'X = X'X$ establishing (i).

(ii) $X'XG'X'X = X'X$

by Lemma 2 $X'XG'X' = X'$

transposing $XGX'X = X$ establishing (ii).

(iii) Suppose F is some generalized inverse different from G. Then $XGX'X = X = XFX'X$. By Lemma 2 $XGX' = XFX'$. That is, XGX' is the same for all generalized inverses of $X'X$, establishing (iii).

D. THE LINEAR MODEL

The general linear model is $Y = Xb + e$ where Y is an $n \times 1$ vector of observations whose components are random and observable; X is an $n \times p$ matrix of experimental design whose components are real and known; b is a $p \times 1$ vector of parameters whose components are real and unknown; e is an $n \times 1$ vector of experimental error whose components are random and unobservable. The vector e is defined as

$$e = Y - E(Y)$$

$$E(e) = E(Y) - E(E(Y)) = 0,$$

and

$$E(Y) = E(Xb) + E(e)$$

$$E(Y) = Xb.$$

Every element in e is assumed to have the same variance v^2

and zero covariance with every other element, thus e is distributed $(0, v^2 I)$ and Y is distributed $(Xb, v^2 I)$. Deriving the normal equations for the linear model yields

$$X'Xb = X'Y$$

which can be solved for b^0 using the techniques of generalized inverses described earlier, i.e.,

$$b^0 = GX'Y$$

and

$$\begin{aligned} E(b^0) &= E(GX'Y) \\ &= GX'E(Y) \\ &= GX'Xb \\ &= Hb \end{aligned}$$

where

$$H = GX'X.$$

III. THE CONCEPT OF ESTIMABILITY

A. ESTIMABILITY

As defined by Searle [1], a linear function q'b of the parameters in b is estimable if it is equal to any linear function [t'E(Y)] of the expected value of the observations in Y. It is important to note that t' is not in general unique; the only requirement for estimability is that such a vector exist.

B. PROPERTIES

The definition of estimability leads to four mathematical properties of immediate importance:

(1) The expected value of any observation is estimable. In this case t' is a vector with a single element equal to one; the rest of its elements are zero.

(2) Any linear combination of estimable functions is estimable. If q'b and r'b are estimable, then q'b = t'E(Y) and r'b = s'E(Y). Therefore $c_1 q'b + c_2 r'b = (c_1 t' + c_2 s')E(Y)$ which is estimable.

(3) An alternative form of the condition of estimability can be developed as follows. If q'b is estimable, then by

definition $q'b = t'E(Y)$ hence $q'b = t'Xb$. This must hold for all values of b since the condition of estimability does not depend on a specific choice of b. This leads to the result $q' = t'X$.

(4) When $q'b^0$ is estimable, $q'b^0$ is invariant to the solution used for b^0 because

$$q'b^0 = t'Xb^0 = t'XGX'Y.$$

Since by Theorem 4, XGX' is invariant to G, $q'b^0$ is invariant to G and therefore to b^0 when $q'b$ is estimable. Herein lies the essential importance of estimability: if $q'b$ is estimable, $q'b^0$ has the same value for all solutions b^0. That is, an estimable function is a linear function of the parameters that is invariant to whatever solution is used for b^0.

C. THE TEST

A function $q'b$ is estimable if there exists some vector t' such that $q' = t'X$. Finding such a vector t' may be a formidable task with a design of large dimensions. As an alternative, it is possible to test for estimability by determining if $q'H = q'$. Searle [1] shows that $q'b$ is estimable if and only if $q'H = q'$, as follows.
If $q'b$ is estimable

$$q' = t'X$$
$$q'H = t'XH$$
$$q'H = t'XGX'X$$

by Theorem 4 GX' is a generalized inverse of X,

hence $$q'H = t'X$$
$$q'H = q'.$$
On the other hand, if
$$q' = q'H,$$
$$q' = q'GX'X$$
and $$q' = t'X \text{ for } t' = q'GX'.$$

D. THE CONSTRAINED MODEL

1. development

The normal equations $X'Xb = X'Y$ form a consistent system of linear equations where X is of rank $r<p$. Because $X'X$ is, in general, not of full rank, there are many solution vectors that will satisfy the system. In order to obtain a particular solution b^0, additional constraints of the form $Cb = 0$ are often added to the model. A commonly used set of constraints satisfies the restrictions

* the main effects sum to zero
* the interaction effects sum to zero across each subscript.

Adding the constraints $Cb = 0$, where the $(p-r)$ rows of C are linearly independent of the rows of X, yields the following system of linear equations:

$$\begin{vmatrix} Y \\ \cdots \\ 0 \end{vmatrix} = \begin{vmatrix} X \\ \cdots \\ C \end{vmatrix} b + \begin{vmatrix} e \\ \cdots \\ 0 \end{vmatrix}$$

The constraint matrix C can be used to transform the design matrix X into a constrained matrix X^* by performing basic row operations on the system of linear equations.

$$\begin{vmatrix} Y \\ \cdots \\ 0 \end{vmatrix} = \begin{vmatrix} X^* \\ \cdots \\ C \end{vmatrix} b + \begin{vmatrix} e \\ \cdots \\ 0 \end{vmatrix}$$

Note that X^* is the same size $(n \times p)$ as X; b remains unchanged. The practical effect of introducing the constraints into the design matrix is to make some of the columns of X^* consist entirely of zeros. While b remains unchanged, the transformation of X into X^* has the effect of "deleting" some elements of the parameter vector by the mechanism of creating those columns of zeros. Once the constraints have been integrated into the design matrix, transforming X into X^*, the constraints become redundant and can be removed from the model by the following technique. Let A be a $(n,n+p-r)$ matrix such that

$A_{ij} = 0$ if $i \neq j$

$A_{ij} = 1$ if $i = j$.

Then multiplying by A,

$$A\begin{vmatrix} Y \\ \cdots \\ 0 \end{vmatrix} = A\begin{vmatrix} X^* \\ \cdots \\ C \end{vmatrix} b + A\begin{vmatrix} e \\ \cdots \\ 0 \end{vmatrix}$$

yields the constrained linear model $Y = X^*b + e$, which is equivalent to the constrained system above.

Since e is assumed normal $(0, v^2 I)$, Y is also normal; $E(Y) = X^*b$. The normal equations, $X^{*\prime}X^*b = X^{*\prime}Y$, can be solved for a particular solution b^0 that will also satisfy the original normal equations $X'Xb = X'Y$. If G^* is defined as the generalized inverse of $X^{*\prime}X^*$ then $b^0 = G^*X^{*\prime}Y$

and it follows that $G^* X^{*\prime}$ is a generalized inverse of X^*. Let $H^* = G^* X^{*\prime} X^*$.

It is stressed that this constrained linear model was developed solely for the purpose of finding some particular solution vector b^0 to the original system of normal equations. In the discussion that follows, X^* is the same size as X and the parameter vector b is the same in the constrained linear model as it was in the original linear model $Y = Xb + e$.

2. <u>example</u>

As a simple example of the development of X^*, assume that the design matrix X is

$$\begin{matrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{matrix}$$

Let C be

$$\begin{matrix} 0 & 1 & 1 & 1. \end{matrix}$$

Then $\begin{vmatrix} X \\ \cdots \\ C \end{vmatrix}$ is

$$\begin{matrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{matrix}$$

Subtracting the bottom row from the top row yields

$$\begin{matrix} 1 & -1 & -1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{matrix}$$

which is $\begin{vmatrix} X^* \\ \cdots \\ C \end{vmatrix}$.

The appropriate A matrix is

$$\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{matrix}$$

Multiplying yields X^*

$$\begin{matrix} 1 & -1 & -1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{matrix}$$

E. BIOMEDICAL COMPUTER PROGRAMS

The University of California publishes and maintains the BIOMED series of standard data analysis packages for use on digital computers [Dixon, 1976]. One of the programs within the package, BMD05V, performs computations for analysis of variance with the linear statistical model. The design matrix employed is not the same as the design matrix (X) in that model however. A user of BMD05V is required to introduce appropriate additional constraints to permit computing a particular solution (b^0) for the parameter vector. It will be shown that techniques applicable to the design matrix X in the general linear model can be applied directly to the BMD05V design matrix.

F. APPLICATION TO BMD05V

The constraints, $Cb = 0$, added to the linear model in section D above are the type used to generate the BMD05V design matrix. The resulting matrix X^* has the same number of columns as the original design matrix X, but because some of these columns are zero, it is possible to suppress them for arithmetic purposes. For computational simplicity, the matrix actually used by the computer program deletes all the zero columns and assumes a corresponding "reparameterized" b vector of lower dimension. For mathematical rigor, however, the X^* used in the following sections retains the same number of columns as X. This restriction will be eased when the matrix X^* is actually applied to the computer programs in Appendix B.

G. ESTIMABILITY IN THE CONSTRAINED MODEL

It can be shown that estimability in the constrained model $Y = X^*b + e$ follows the same pattern as estimability in the full model $Y = Xb + e$.

Theorem 5. $q'b$ is estimable if and only if $q'H^* = q'$.
Proof: By definition, $q'b$ is estimable if

$$q'b = t'E(Y), \text{ i.e., if}$$

$$q'b = t'X^* b, \text{ i.e., if}$$

$$q' = t'X^* . \text{ Then}$$

$$q'H^* = t'X^* G^* X^* 'X^* ,$$

$$q'H^* = t'X^* = q',$$

and if $$q' = q'H^* ,$$

$$q' = q'G^* X^* 'X^* . \quad \text{Let } t' = q'G^* X^* '.$$

Then $$q' = t'X^* .$$

This result allows computations to be performed directly upon the constrained matrix in order to examine the estimability of proposed hypotheses. The computer program HYTEST (Appendix A) can accept either the constrained design matrix X^* (with "zero" columns suppressed) or the standard design matrix X as an input. If X is used for input, HYTEST offers the option of using either the constrained matrix X^* or the standard matrix X to compute tests of estimability.

Note that $q'H^* b$ is always estimable since $q'H^* b = q'G^* X^* 'X^* b = q'G^* X^* 'E(Y) = t'E(Y)$ where $t' = q'G^* X^* '$.

H. TESTABILITY

1. the hypothesis

From Searle [1], all linear hypotheses can be handled by a general procedure; specific hypotheses are then

considered to be applications of this general procedure.

The general hypothesis may be written $H_0: K'b = m$ where K' is a matrix of s rows and p columns. The only limitation on K' is that it have full row rank. That is, the hypothesis must be composed of linearly independent functions of the parameter vector.

2. <u>analysis of variance</u>

To review analysis of variance briefly, classical techniques rely upon the ratio of two independent Chi-square distributions, each divided by its respective degrees of freedom, to generate an F statistic. The sum of squares explained by the model if the hypothesis is assumed true, divided by its degrees of freedom forms the numerator of the statistic. For many situations, the denominator is the sum of squares for error divided by its degrees of freedom. Each sum of squares can be conveniently represented by appropriate quadratic forms which must meet certain requirements in order to be Chi-square distributed.

Searle's derivation of a test of the general hypothesis depends upon K'b being estimable for every row $k_i'b$. If this assumption is satisfied, the quadratic form

$$Q/v^2 = (K'b^0 - m)'(K'GK)^{-1}(K'b^0 - m)/v^2$$

is distributed non-central Chi-square and has rank s. The sum of squares for error can be shown to be

$$SSE = (Y-XK(K'K)^{-1}m)'(I-XGX')(Y-XK(K'K)^{-1}m).$$

Q and SSE are independent so $F(H) = Q/s/SSE/(n-r)$ is distributed non-central

$$F'[s, n-r, (K'b-m)'(K'GK)^{-1}(K'b-m)/2v^2]$$

The test statistic is

$$F(H) = (K'b^0 - m)'(K'GK)^{-1}(K'b^0 - m)/s/SSE/(n-r)$$

which is F distributed with s and n-r degrees of freedom under the null hypothesis.

3. <u>determination of testability</u>

Suppose that $K'b$ is not estimable. Then the hypothesis H_0: $K'b = m$ is not testable. Assuming that $(K'GK)^{-1}$ exists, if one were to compute $F(H)$, what hypothesis is actually being tested? When working with the constrained linear model $Y = X^*b + e$, the answer is "$K'H^*b = m$." The derivation which follows closely parallels the procedure used by Searle [1] for the linear model $Y = Xb + e$.

The hypothesis H: $K'H^*b = m$ is testable since $K'H^*b$ is estimable for each row k_i'. The appropriate numerator quadratic form is

$$Q_1 = (K'H^*b^0 - m)'(K'H^*G^*H^*'K)^{-1}(K'H^*b^0 - m)$$

But $K'H^*b^0 = K'G^*X^*'X^*G^*X^*'Y$ and since

$X^*G^*X^*' = X^*G^*'X^*'$, it follows

$K'H^*b^0 = K'G^*X^*'X^*G^*'X^*'Y$.

Let $G_1^* = G^*X^*'X^*G^*'$ (which is a generalized inverse of

$X^{*\prime}X^{*}$). Then

$$K'H^{*}b^{0} = K'G_1 X^{*\prime}Y$$

$$K'H^{*}b^{0} = K'b_1^{0}$$

where $b_1^{0} = G_1 X^{*\prime}Y$ is the solution to $X^{*\prime}X^{*}b^{0} = X^{*\prime}Y$ obtained from using G_1^{*}.

Also

$$K'H^{*}G^{*}H^{*\prime}K = K'G^{*}X^{*\prime}X^{*}G^{*}X^{*\prime}X^{*}G^{*\prime}K$$

$$K'H^{*}G^{*}H^{*\prime}K = K'G^{*}X^{*\prime}X^{*}G^{*\prime}K$$

$$K'H^{*}G^{*}H^{*\prime}K = K'G_1^{*}K$$

Therefore Q_1 reduces to

$$Q_1 = (K'b_1^{0} - m)'(K'G_1^{*}K)^{-1}(K'b_1^{0} - m)$$

Which is the quadratic form that would result from attempting to test the non-testable hypothesis $K'b = m$ using the solution $b_1^{0} = G_1 X^{*\prime}Y$. These calculations are indistinguishable from those that would be performed in testing the testable hypothesis $K'H^{*}b = m$.

IV. SYNOPSIS OF RESULTS

A. CONCLUSIONS

1. A given linear function of the parameter vector ($q'b$) is estimable if and only if $q'H = q'$.

2. Since only estimable functions are testable, if $q'H \neq q'$ then the hypothesis actually being tested is not H_0: $q'b = m$ but rather H_0: $q'Hb = m$.

3. The mathematics developed for proving both of the preceding conclusions can also be applied to a constrained design matrix, such as that used in the BMD05V program, to allow determination of estimability directly. That is,

- the function $q'b$ is estimable if and only if $q'H^* = q'$, and
- if $q'b$ is not estimable then the hypothesis H_0: $q'b = m$ is actually testing H_0: $q'H^*b = m$.

B. SIGNIFICANCE

The results presented above afford a mathematical justification of the need for a computer program to

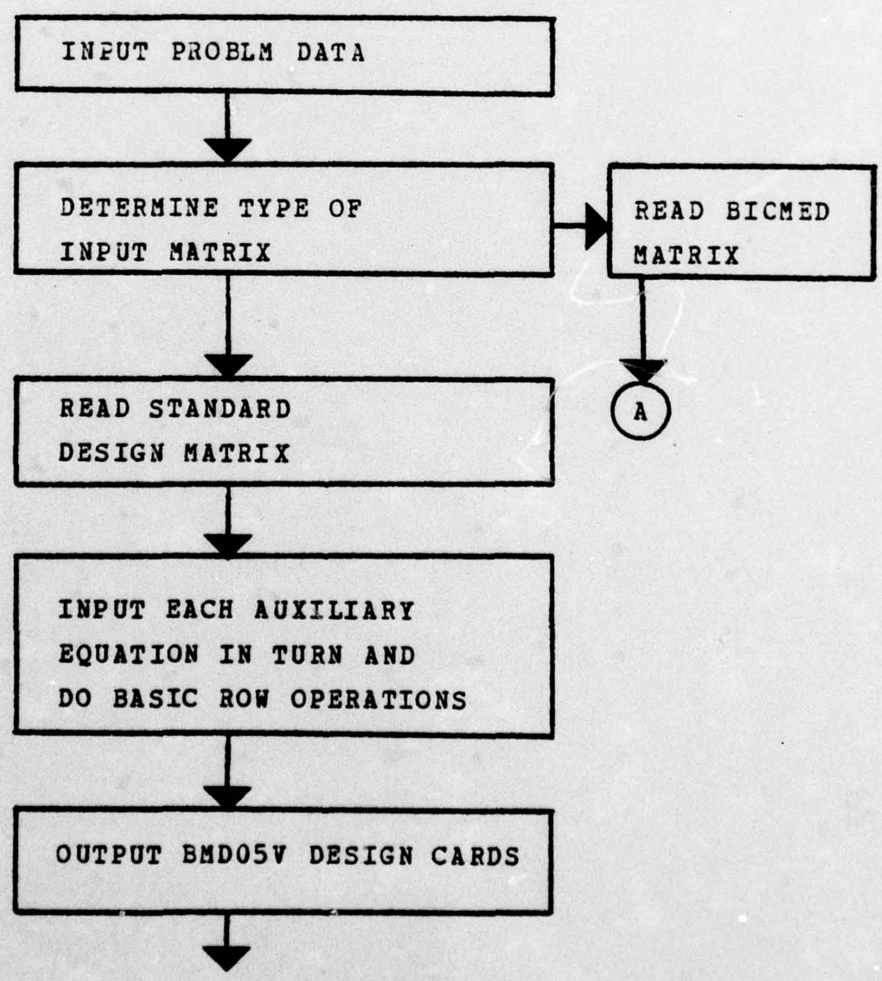
determine the testability of proposed hypotheses. A sample program, HYTEST, is presented in Appendix D. Using the HYTEST program permits an analyst to determine not only whether each of his hypotheses is testable, but also precisely what main effects and interactions confound each hypothesis that is not testable. Such information can be used to make more informed decisions, a priori, on the design of experiments, and a posteriori on the analysis and interpretation of experimental results. Knowing the nature and degree of confounding may not necessarily ease decision making. It can help to ensure that once a hypothesis is accepted or rejected the analyst is aware of the degree of "purity" of his conclusions concerning the effects of various factors in the experiment. If used, these techniques can prevent an analyst from complacently assuming that he is testing one hypothesis when he is, in fact, testing something quite different.

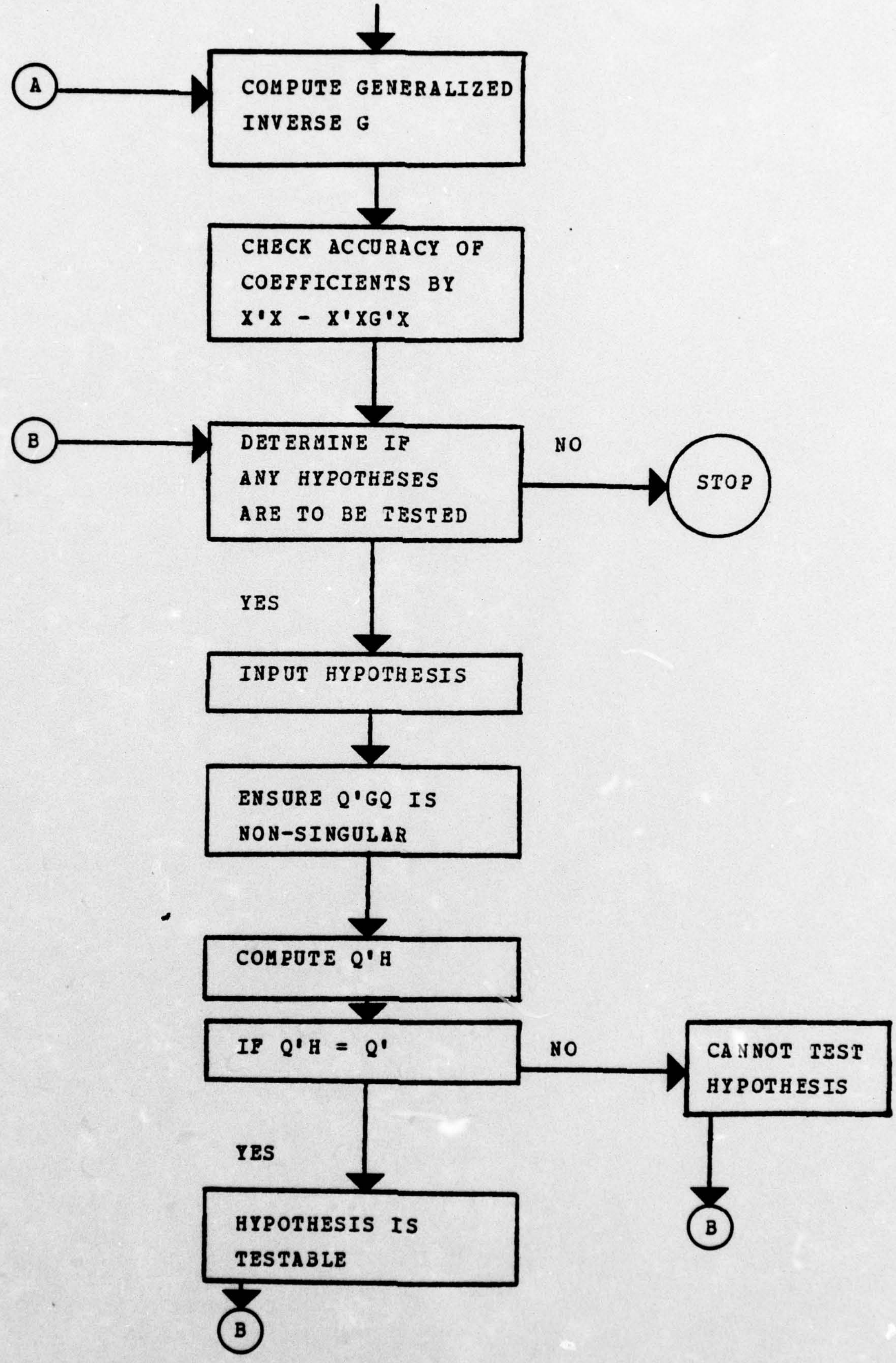
APPENDIX A

HYTEST

HYTEST is a FORTRAN IV program that facilitates determining if a proposed hypothesis is testable. The User's Guide contains operating instructions and a list of options.

FLOWCHART




A
COMPUTE GENERALIZED INVERSE G
CHECK ACCURACY OF COEFFICIENTS BY X'X - X'XG'X
B
DETERMINE IF ANY HYPOTHESES ARE TO BE TESTED
NO
STOP
YES
INPUT HYPOTHESIS
ENSURE Q'GQ IS NON-SINGULAR
COMPUTE Q'H
IF Q'H = Q'
NO
CANNOT TEST HYPOTHESIS
B
YES
HYPOTHESIS IS TESTABLE
B

SUBROUTINE

In the interest of computational accuracy and speed, HYTEST uses subroutine LPSDOR from the International Mathematical and Statistical Library (IMSL) to compute a generalized inverse G. This is a proprietary subroutine. Under the licensing agreement, its code may not be distributed to or used by a user outside the Naval Postgraduate School. It may not be used on a non-NPS computer system.

APPENDIX B

USER'S GUIDE

This guide contains complete operating instructions for using the ccmputer program HYTEST.

the constrained design matrix X^* actually used in the computer programs BMD05V and HYTEST is the same X^* introduced in Chapter III except that all zero columns have been deleted to facilitate computation. This has the effect of reducing the parameter vector b; an analyst using these programs is cautioned to ensure that he is aware of which elements have been "deleted" from the parameter vector. See Chapter III, sections D, E and F, for further information.

A. OVERVIEW OF COMPUTER PROGRAM HYTEST

The program HYTEST exploits techniques enumerated earlier in this paper to determine if selected hypotheses of the form H_0: $k_i'b = m$ are testable within the framework of a specified linear model. If an hypothesis is not testable, HYTEST computes an algebraic form of the parameter vector that would actually be tested if the proposed hypothesis were to be employed.

Given the standard design matrix (X) from the linear model, and appropriate additional constraints, HYTEST can compute the design matrix (X^*) for use in the BMD05V program. If desired, DESIGN cards for use in the BIOMED package can be produced as an auxiliary output. If the BIOMED option is exercised, HYTEST performs all of its calculations upon X^* making it essential for the user to be aware of the exact structure not only of the original design matrix X, but also of the constrained matrix X^*. The user can select the option to have all calculations performed on X, if preferred.

B. INPUT OPTIONS

The program can accept either the standard design matrix from the basic linear model, or a BMD05V design matrix.

C. OUTPUT OPTIONS

1. hypotheses

The major option for output concerns the testability of user defined hypotheses. For any number of hypotheses from zero to 99, HYTEST will determine if each hypothesis is testable, and if not testable, the program computes what would actually be tested if the specified hypothesis were to be employed. This option is automaticllly suppresssed if there are no hypotheses to be tested.

2. design cards

If a BIOMED design matrix is computed, the user can select an option that will punch appropriate DESIGN cards for BMD05V; a printed replica of the cards is also produced. If punched cards are not required, the user can opt for the printed form of the DESIGN cards without having them punched. This output option is only available if a standard design matrix X is used for input.

3. accuracy

An estimate of the accuracy with which HYTEST is computing the generalized inverse matrix G can be obtained upon request. The output consists of the matrix resulting from subtracting X'X - X'XGX'X. If the computer were perfectly accurate, all entries would be zero. Because of arithmetic inaccuracies in computing G, the entries are frequently not zero. Since the matrix H=GX'X is used in assessing the testability of hypotheses, X'X -X'XGX'X affords an estimate of the accuracy to be expected when determining the nature of the hypothesis being tested.

4. H = GX'X

The H matrix used in determining the testability of selected hypotheses will be printed if requested. Selecting this option will generate printed output for all of the previous options as well.

5. generalized inverse G

The final option prints out the generalized inverse of X'X used in computations within HYTEST. This option automatically includes printed output for all previous options as well.

D. INPUT REQUIREMENTS

1. card order

Input cards must be in the following order:

PROBLM card (required)
Design matrix cards (required)
AUXEQN card (required if standard
   design matrix input option is used)
Auxiliary equations (optional)
Hypotheses (optional)
FINISH card (required after last problem)

2. problem card (required)

The HYTEST PROBLM card, based upon a similar card used in the BMD05V program, is used to set up various program parameters and options.

DATA	COLUMNS	RESTRICTIONS
PROBLM	1-6	
User's optional problem number	7-8	

Number of design card sets	9-11	$1 \le ND \le 150$
Number of columns in design matrix	12-13	$1 \le NC \le 60$
Blank	14-15	
Number of hypotheses	16-17	$1 \le NH \le 99$
Blank	18-26	
Output options	27	
Blank	28-71	
Input options	72	0= X matrix 1= X^* matrix

Output options are selected from the following table.

IF COLUMN 27 CONTAINS	THE PRINTED OUTPUT WILL INCLUDE
0	A. An assessment of what each hypothesis is testing
1	B. Design cards (if the BIOMED output option is being used) plus option 0
	C. Accuracy of coefficients and option 0
2	D. The H matrix (H=GX'X); plus all output from options 0 and 1
3	E. The generalized inverse G plus all previous options.

3. <u>design matrix cards (required)</u>

a. standard design matrix

One design card is required for each unique row of the design matrix. The first 2 columns of card i contain (right justified) the number of rows in the design matrix that are identical to row i. The next 60 columns (3-62) are reserved for the columns of the design matrix. Enter a zero

or one in the appropriate card column. Each card column corresponds to a column in the design matrix (not to exceed a total of 60 columns).

example

If a hypothetical design matrix were:

```
1 1 0 0
1 1 0 0
1 0 1 0
1 0 1 0
1 0 1 0
1 0 0 1
```

Only three design cards are required:

```
021100   (in columns 1-6)
031010
011001
```

b. BMD05V design matrix

Enter the BIOMED design matrix (without data cards) exactly as it is used in the BMD05V package.

4. <u>AUXEQN card (required if the standard design matrix input is used)</u>

DATA	COLUMNS	RESTRICTIONS
AUXEQN	1-6	
Number of auxiliary equations	7-8	$0 \leq NA \leq 99$
Punched card output option	9-11	1=Design cards Punched 0=Design cards Not punched

Caution: Do not use this card if the BIOMED input option is exercised.

If there are no auxiliary equations, the program will perform its computations on the standard design matrix. None of the BIOMED output options are then available.

5. Auxiliary equations (optional)

Auxiliary equations are algebraic equivalents of BMD05V constraints. For the example to be used in this section assume the linear model

$$Y_{ij} = u + a_i + e_{ij} \qquad i=1,2,3$$

Each auxiliary equation (m) requires a separate series of input cards. The first card contains the number of parameters in equation m. As an example, the auxiliary equation

$$a_1 + a_2 + a_3 = 0$$

contains three parameters so the first card for this equation would contain 03 in the first two columns.

The second card contains the column number and the coefficient of the parameter whose column is to be deleted from the BMD05V matrix. For instance, the parameter vector in this example is

$$b' = (u, a_1, a_2, a_3)$$

Parameter a_3 corresponds to the fourth column of the design matrix X. In order to eliminate a_3 from the BMD05V matrix, the second card must contain:

0401

in the first four columns.

A separate card is required for each parameter in the auxiliary equation, so this example needs two more cards:

0201

0301

corresponding to the columns and coefficients for a_1 and a_2 respectively.

Caution: if a parameter's column is to be deleted from the BMD05V design matrix, that parameter must not be used in an auxiliary equation; its algebraic equivalent must be used instead. A complete set of cards must be included for each separate auxiliary equation (m cards for each auxiliary equation with m parameters).

The above input will produce the following BMD05V design cards:

```
DESIGN002001001000
DESIGN003001000001
DESIGN0010010-10-1
```

6. <u>Hypothesis cards (optional)</u>

The format for hypothesis cards is similar to that for the auxiliary equation cards. Assume that the hypothesis of interest is:

$$H_0: a_1 - a_2 = 0$$

There are two parameters in the hypothesis (a_1 and a_2) corresponding to columns 2 and 3 in the constrained BMD05V matrix. It is important to note that the parameters used in hypothesis testing must be associated with the matrix used to compute H = GX'X. If the standard design matrix $\dot{X}$ is used throughout the program, the column corresponding to a specific parameter is unchanged from the original model formulation. If, on the other hand, the standard design matrix X is used for input, and auxiliary

equations are used to reduce X to the BMD05V matrix X^*, then the columns used in the hypothesis cards must be the appropriate columns from X^*. The first card, for this example hypothesis, contains 02 (for two parameters) in the first two columns. The second and third cards contain

021.0

03-1.0

respectively. The first two columns identify the parameter's column in the appropriate design matrix. The parameter's coefficient, a decimal point and, if appropriate, a minus sign, must be punched in columns 3 through 11.

7. FINISH card (Required after last problem)

If several problems are to be run in sequence, the cards for each problem are to be grouped in sequential blocks. Once HYTEST finishes a problem it determines if another problem is to be run. If so, it executes that problem. It will continue executing problems in sequence until all problems are completed and a FINISH card is encountered. The FINISH card must be the last card in the data deck and the word FINISH must be punched in the first six columns.

APPENDIX C

EXAMPLE

A. THE MODEL

The use of HYTEST can best be illustrated through use of an example. The one chosen is from Searle [1].

The model is:

$$Y_{ijk} = m + A_i + B_j + AB_{ij} + e_{ijk} \qquad \begin{array}{l} i=1,2,3 \\ j=1,\ldots,4 \\ k=1,\ldots,4 \end{array}$$

Y_{ijk} is the kth observation of the ith treatment of the jth type; M is the mean effect; A_i is the effect of the ith treatment; B_j is the effect of the jth type; AB_{ij} is the interaction between the ith treatment and the jth type and e_{ijk} is the error term. The number of observations noted for each cell is shown in the following table:

	j=1	j=2	j=3	j=4
i=1	3	0	1	2
i=2	2	2	0	0
i=3	0	2	2	4

The X matrix is:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
1	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
1	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1

Suppressing the zero columns for AB_{12}, AB_{23}, AB_{24}, and AB_{31} the input design matrix for HYTEST is

```
031100100010000000
011100001001000000
021100000100100000
021010100000010000
021010010000001000
021001010000000100
021001001000000010
041001000100000001
```

B. CARD PREPARATION

The size of the BMD05V matrix is controlled by the type of constraints used in the BIOMED package. In general, the number of parameters for each main effect and each interaction effect is reduced by 1 in each dimension. In this example, 1 parameter is deleted for each of the main effects, A and B. Since $i = 1,2,3$ and $j = 1,2,3,4$ there are $3 \times 4 = 12$ interaction terms. Reducing each dimension by 1 yields $i = 1,2$ and $j = 1,2,3$ for $2 \times 3 = 6$ remaining interaction terms. As noted in the preparation of the input design matrix above, the columns for four interaction terms are zero. Those four terms are therefore deleted from the model. Two other interaction and two main effect terms can be deleted by addition of the usual BMD05V constraints. For this example, the following constraints were adopted:

(1) $$A_1 + A_2 + A_3 = 0$$

(2) $$B_1 + B_2 + B_3 + B_4 = 0$$

(3) $$AB_{11} + AB_{12} + AB_{13} + AB_{14} = 0$$

(4) $$AB_{31} + AB_{32} + AB_{33} + AB_{34} = 0$$

In auxiliary equation (1) it is desired to delete parameter A_3 which corresponds to column 4 of the design matrix. Since there are three parameters in this equation the first card is:

03

The other cards are:

0401

0201

0301

This completes the cards for auxiliary equation (1).

Auxiliary equation (2) is quite similar. If B_4 is chosen for deletion the cards are:

04
0801
0701
0601
0501

Auxiliary equation (3) contains the parameter AB_{12} which does not appear in the design matrix (it is a column of 0's). Under the BMD05V constraints, it has an algebraic equivalent, $AB_{12} = -AB_{22} - AB_{32}$, which must be used in its place. Equation (3) then becomes

(3a) $$AB_{11} - AB_{22} - AB_{32} + AB_{13} + AB_{14} = 0.$$

The input cards used to delete AB_{14} from the design matrix are:

05
1101
13-1
14-1
1001
0911

The interaction term AB_{31} from equation (4), does not appear in the design matrix. Substituting $AB_{31} = - AB_{11} - AB_{21}$ into equation (4) yields:

(4a) $$-AB_{11} - AB_{21} + AB_{32} + AB_{33} + AB_{34} = 0.$$

The input cards used to delete AB_{34} from the design matrix are:

05
1601
09-1
12-1
1401
1501

Deleting a total of eight parameters from the standard design matrix will yield a BMD05V matrix of twelve non-zero columns.

1	2	3	4	5	6	7	8	9	10	11	12
M	A_1	A_2	B_1	B_2	B_3	AB_{11}	AB_{13}	AB_{21}	AB_{22}	AB_{32}	AB_{33}

C. HYPOTHESES

To verify that $A_1 - A_2 + AB_{11} - AB_{21}$ is estimable but that $A_1 - A_2$ is not, the following cards are used:

04
021.0
03-1.0
071.0
09-1.0
02
021.0
03-1.0

After completing the PROBLM card, the AUXEQN card and the FINISH card as outlined in the User's Guide and putting the cards in the correct order, one is ready to run HYTEST.

The outputs of major interest are the DESIGN cards and the hypotheses. In the interest of brevity, only those two outputs will be illustrated.

D. SAMPLE OUTPUT

Design cards suitable for BMD05V will be similar to the sample shown below.

```
DESIGN  3  1  1  0  1  0  0  1  0  0  0  0  0
DESIGN  1  1  1  0  0  0  1  0  1  0  0  0  0
DESIGN  2  1  1  0 -1 -1 -1 -1 -1  0  1  1  0
DESIGN  2  1  0  1  1  0  0  0  0  1  0  0  0
DESIGN  2  1  0  1  0  1  0  0  0  0  1  0  0
DESIGN  2  1 -1 -1  0  1  0  0  0  0  0  1  0
DESIGN  2  1 -1 -1  0  0  1  0  0  0  0  0  1
DESIGN  4  1 -1 -1 -1 -1 -1  1  0  1  0 -1 -1
```

The column on the left below represents the vector q' in the hypothesis H: q'b = m. The column on the right below represents the vector q'h. If q' = q'H, the hypothesis is testable. If q' ≠ q'H, the hypothesis is not testable.

THE HYPOTHESIS OF INTEREST IS (HO: Q'B = M)	THE HYPOTHESIS WHICH IS ACTUALLY BEING TESTED (HO: Q'HB = M)
0.0	0.115
1.0	0.748
-1.0	-0.656
0.0	-0.084
0.0	-0.298
0.0	0.038
0.0	0.221
0.0	0.099

0.0	-0.374
0.0	-0.160
0.0	0.275
0.0	-0.061

WITHIN THE LIMITS OF COMPUTATIONAL ACCURACY,
HYPOTHESIS 1 CANNOT BE TESTED AS STATED ABOVE

THE HYPOTHESIS OF INTEREST IS (HO: Q'B = M)	THE HYPOTHESIS WHICH IS ACTUALLY BEING TESTED (HO: Q'HB = M)
0.0	0.0
1.0	1.0
-1.0	-1.0
0.0	0.0
0.0	0.0
0.0	0.0
1.0	1.0
0.0	0.0
-1.0	-1.0
0.0	0.0
0.0	0.0
0.0	0.0
0.0	0.0

WITHIN THE LIMITS OF COMPUTATIONAL ACCURACY,
HYPOTHESIS 2 CAN BE TESTED AS STATED ABOVE

As expected (Searle [1]), one hypothesis is estimable; the other is not. Additional output information is available from HYTEST if desired. The User's Guide enumerates all of the options.

APPENDIX D

COMPUTER PROGRAM

```
      REAL*8 BP,BC,BF
      DIMENSION X(150,60),NREPS(150),XS(150,60)
      DIMENSION XR(60),ID(60),XSX(60,60),WKAREA(3840),Q(60)
      DIMENSION GH(60), H(60,60)
      DATA BP/'PROBLM  '/,BF/'FINISH  '/
1000  FORMAT ('0','PROBLEM NUMBER',I4)
1001  FORMAT ('0','NUMBER OF DESIGN CARD SETS',I6)
1002  FORMAT ('0','NUMBER OF INDEPENDENT VARIABLES IN DESIGN
     1 MATRIX',I5)
1003  FORMAT ('0','NUMBER OF INDEPENDENT VARIABLES IN BMD05V
     1 MATRIX',I5)
1010  FORMAT ('0','WITHIN THE LIMITS OF COMPUTATIONAL ACCURA
     1CY,')
1011  FORMAT (' ','HYPOTHESIS',I3,2X,'CAN BE TESTED AS STATE
     1D ABOVE')
1012  FORMAT (' ','HYPOTHESIS',I3,2X,'CANNOT BE TESTED AS ST
     1ATED ABOVE')
1026  FORMAT ('0','GENERALIZED INVERSE MATRIX, G'/)
1027  FORMAT ('0','H MATRIX (GX"X)'/)
1030  FORMAT (' ',25F5.1)
1035  FORMAT ('0','ACCURACY OF COEFFICIENTS (X"X - X"XGX"X)'
     1/)
1036  FORMAT (' ',10(3X,F10.5))
1037  FORMAT ('0')
1040  FORMAT ('0','INPUT OPTION: STANDARD DESIGN MATRIX')
1041  FORMAT ('0','INPUT OPTION: BMD05V DESIGN MATRIX')
1050  FORMAT (A6,I2,I3,I2,2X,I2,8X,I2,44X,I1)
1051  FORMAT ('0','FINISH CARD MISSING OR NEXT PRCBLM CARD O
     1UT OF SEQUENCE. PROGRAM CANNOT CONTINUE AND IS TERMINA
     2TING.')
1060  FORMAT ('0','THE HYPOTHESIS',10X,'THE HYPOTHESIS WHICH
     1 IS')
1061  FORMAT (' ',5X,F7.3,18X,F7.3)
1062  FORMAT (' ','OF INTEREST IS',10X,'ACTUALLY BEING TESTE
     1D')
1063  FORMAT (' ','(HO:   Q"B = M)',10X,'(HO:   Q"HB = M)'//)
1066  FORMAT ('0','HYPOTHESIS',I3,' CANNOT BE TESTED BECAUSE
     1 G"GQ IS SINGULAR')
1070  FORMAT (6X,I2,I3,I2,2X,I2,8X,I2,44X,I1)
1071  FORMAT (I2,60F1.0)
1072  FORMAT (6X,I3,21F3.0/6X,22F3.0)
1075  FORMAT (I2,30F2.0)
1076  FORMAT (' ','DESIGN',I3,21I3 / 7X,22I3)
1077  FORMAT ('0','DESIGN CARDS'/)
1078  FORMAT ('DESIGN',I3,21I3/'DESIGN',22I3/'DESIGN',22I3)
1079  FORMAT (I2,F9.0)
1090  FORMAT ('0','DESIGN CARD NUMBER',I4,2X,'IS  INVALID')
C READ IN THE DESIGN MATRIX AND THE PROBLEM CARD
      READ (5,1070) NP,ND,N,NH,IPR,IO
2270  WRITE (6,1000) NP
      WRITE (6,1001) ND
      WRITE (6,1002) N
      IF (IO .EQ. 0) GO TO 2305
C THIS READS IN THE BIOMED DESIGN MATRIX
      WRITE (6,1041)
      DO 2300 I=1,ND
```

```
      READ (5,1072) NREPS(I), (XS(I,J),J=1,N)
      IF (NREPS (I) .LE. 0) GO TO 2250
 2300 CONTINUE
      NS=N
      GO TO 2365
C THIS READS IN THE STANDARD DESIGN MATRIX
 2305 WRITE (6,1040)
      DO 2310 I=1,ND
      READ (5,1071) NREPS (I ), (X(I ,JR),JR=1,N)
      IF (NREPS (I) .LE. 0) GO TO 2250
 2310 CONTINUE
C COMPUTE THE BMD05V MATRIX (X*)
      READ (5,1070) NA,IPCH
      NS=N-NA
      IF (NA .NE. 0) GO TO 2315
      DO 2313 I=1,ND
      DO 2313 J=1,NS
      XS(I,J)=X(I,J)
 2313 CONTINUE
      GO TO 2365
 2315 WRITE (6,1003) NS
C READ IN EACH AUXILIARY EQUATION IN TURN
C AND DO THE BASIC ROW OPERATIONS
      DO 2330 IS=1,NA
      DO 2320 IR=1,N
      XR(IR)=0.0
 2320 CONTINUE
      READ (5,1075) NVAL
      NVAL=NVAL-1
      READ (5,1075) ID(IS),XVAL
      XR(ID(IS))=XVAL
      DO 2323 IV=1,NVAL
      READ (5,1075) NL,XVAL
      XR(NL)=XVAL
 2323 CONTINUE
      DO 2325 IC=1,ND
      IF(X(IC,ID(IS)) .EQ. 0.0) GO TO 2325
      DO 2324 JX=1,N
      X(IC,JX)=X(IC,JX)-XR(JX)
 2324 CONTINUE
 2325 CONTINUE
 2330 CONTINUE
C SUPPRESS ZERO ROWS AND COLUMNS
      JR=1
      DO 2340 JT=1,N
      DO 2335 JX=1,NA
      IF (ID(JX) .EQ. JT) GO TO 2340
 2335 CONTINUE
      DO 2336 IS=1,ND
      XS(IS,JR)=X(IS,JT)
 2336 CONTINUE
      JR=JR+1
 2340 CONTINUE
C OUTPUT BMD05V DESIGN CARDS
      IF (IPR .LT. 1) GO TO 2365
      WRITE (6,1077)
      DO 2360 J=1,ND
      DO 2355 K=1,NS
      ID(K)=0
      IF (XS(J,K) .LE. -0.5) ID(K)= -1
      IF (XS(J,K) .GE.  0.5) ID(K)= 1
 2355 CONTINUE
      WRITE (6,1076) NREPS (J),(ID(K),K=1,NS)
      IF (IPCH .LT. 1) GO TO 2360
      WRITE (7,1078) NREPS(J), (ID(K),K=1,NS)
 2360 CONTINUE
C COMPUTE X'X=XS
 2365 DO 2380 I=1,NS
      DO 2380 K=1,NS
      XSX(I,K)=0.0
      DO 2380 J=1,ND
      NFIX=NREPS(J)
```

```
      DO 2380 JFIX=1,NFIX
      XSX(I,K)=XS(J,I)*XS(J,K)+XSX(I,K)
2380  CONTINUE
      CALL LPSDOR (XSX,NS,NS,60, H,5,WKAREA,IER)
C TRANSPOSE THE PSEUDO-INVERSE TO GET THE GENERALIZED
C INVERSE (IN X)
      DO 2390 J=1,NS
      DO 2390 I=1,NS
      X(J,I)= H(I,J)
2390  CONTINUE
      IF (IPR .LT. 3) GO TO 2395
      WRITE (6,1026)
      NW=1
2392  NEND=NW+9
      IF (NEND .GT. NS) NEND=NS
      DO 2393 IW=1,NS
      WRITE (6,1036) (X(IW,JW),JW=NW,NEND)
2393  CONTINUE
      WRITE (6,1037)
      NW=NW+10
      IF (NEND .LT. NS) GO TO 2392
C COMPUTE H = GX'X
2395  DO 2400 I=1,NS
      DO 2400 K=1,NS
      H(I,K)=0.0
      DO 2400 J=1,NS
      H(I,K)=X(I,J)*XSX(J,K)+H(I,K)
2400  CONTINUE
      IF (IPR .LT. 2) GO TO 2405
      WRITE (6,1027)
      NW=1
2402  NEND=NW+9
      IF (NEND .GT. NS) NEND=NS
      DO 2403 IW=1,NS
      WRITE (6,1036) (H(IW,JW),JW=NW,NEND)
2403  CONTINUE
      WRITE (6,1037)
      NW=NW+10
      IF (NEND .LT. NS) GO TO 2402
C CHECK H BY COMPUTING X'XGX'X
2405  IF (IPR .LT. 1) GO TO 2422
      DO 2410 I=1,NS
      DO 2410 K=1,NS
      XS(I,K)=0.0
      DO 2410 J=1,NS
      XS(I,K)=XSX(I,J)*H(J,K)+XS(I,K)
2410  CONTINUE
      DO 2415 I=1,NS
      DO 2415 J=1,NS
      XS(I,J)=XS(I,J)-XSX(I,J)
2415  CONTINUE
      WRITE (6,1035)
      NW=1
2418  NEND=NW+9
      IF (NEND .GT. NS) NEND=NS
      DO 2420 IW=1,NS
      WRITE (6,1036) (XS(IW,JW),JW=NW,NEND)
2420  CONTINUE
      WRITE (6,1037)
      NW=NW+10
      IF (NEND .LT. NS) GO TO 2418
C INPUT THE HYPOTHESES TO BE TESTED
2422  IF (NH .EQ. 0) GO TO 9999
      DO 2200 LOOP=1,NH
      DO 2423 I=1,NS
      Q(I)=0.0
2423  CONTINUE
      READ (5,1075) NHYP
      DO 2425 I=1,NHYP
      READ (5,1079) NQ,HVAL
      Q(NQ)=HVAL
2425  CONTINUE
```

```
C CHECK TO ENSURE THAT Q'GQ IS NONSINGULAR
      CHECK=0.0001
      DO 2430 K=1,NS
      XR(K)=0.0
      DO 2430 J=1,NS
      XR(K)=Q(J)*X(J,K)+XR(K)
2430  CONTINUE
      HOLD=0.0
      DO 2440 K=1,NS
      HOLD=XR(K)*Q(K)+HOLD
2440  CONTINUE
      IF (ABS(HOLD) .GT. CHECK) GO TO 2450
      WRITE (6,1066) LOOP
      GO TO 2200
2450  WRITE (6,1060)
      WRITE (6,1062)
      WRITE (6,1063)
C MULTIPLY Q'H
      DO 2210 K=1,NS
      QH(K)=0.0
      DO 2210 J=1,NS
      QH(K)=Q(J)*H(J,K)+QH(K)
2210  CONTINUE
      DO 2215 IW=1,NS
      WRITE (6,1061) Q(IW), QH(IW)
2215  CONTINUE
      WRITE (6,1010)
      DO 2460 K=1,NS
      QA=ABS(Q(K))
      QHA=ABS(QH(K))
      IF (QHA .GT. QA+CHECK .OR. QHA .LT. QA-CHECK)
     1 GO TO 2470
2460  CONTINUE
      WRITE (6,1011) LOOP
      GO TO 2200
2470  WRITE (6,1012) LOOP
2200  CONTINUE
      GO TO 9999
2250  WRITE (6,1090) I
9999  READ (5,1050) BC,NP,ND,N,NH,IPR,IC
      IF (BC .EQ. BP) GO TO 2270
      IF (BC .EQ. BF) GO TO 9000
      WRITE (6,1051)
9000  STOP
      END
```

LIST OF REFERENCES

1. Searle, S. R., *Linear Models*, p. 1 to 316, Wiley, 1971.

2. Dixon, W. J. (ed.), *Biomedical Computer Programs*, p. 615 to 732, University of California Press, 1976.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	2
4.	Professor D. R. Barr, Code 55Bn Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
5.	Professor F. R. Richards, Code 55Rh Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
6.	LCDR William H. Walls, USN 586-B Sampson Lane Monterey, California 93940	1